

# *An LDA and probability-based classifier for the diagnosis of Alzheimer's Disease from structural MRI*

Conference or Workshop Item

Accepted Version

Spedding, A. L., Di Fatta, G. and Saddy, J. D. (2015) An LDA and probability-based classifier for the diagnosis of Alzheimer's Disease from structural MRI. In: The IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 9-12 Nov 2015, Washington D.C., pp. 1404-1411. doi: <https://doi.org/10.1109/BIBM.2015.7359883> Available at <https://centaur.reading.ac.uk/50927/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://dx.doi.org/10.1109/BIBM.2015.7359883>

To link to this article DOI: <http://dx.doi.org/10.1109/BIBM.2015.7359883>

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# An LDA and Probability-based Classifier for the Diagnosis of Alzheimer's Disease from Structural MRI

Alexander Luke Spedding\*, Giuseppe Di Fatta\*, James Douglas Saddy†, and the Alzheimer's Disease Neuroimaging Initiative

\*School of Systems Engineering, University of Reading, UK  
Email: a.l.spedding@pgr.reading.ac.uk, g.difatta@reading.ac.uk

†Centre for Integrative Neuroscience & Neurodynamics, University of Reading, UK  
Email: j.d.saddy@reading.ac.uk

**Abstract**—In this paper a custom classification algorithm based on linear discriminant analysis and probability-based weights is implemented and applied to the hippocampus measurements of structural magnetic resonance images from healthy subjects and Alzheimer's Disease sufferers; and then attempts to diagnose them as accurately as possible. The classifier works by classifying each measurement of a hippocampal volume as healthy control-sized or Alzheimer's Disease-sized, these new features are then weighted and used to classify the subject as a healthy control or suffering from Alzheimer's Disease. The preliminary results obtained reach an accuracy of 85.8% and this is a similar accuracy to state-of-the-art methods such as a Naive Bayes classifier and a Support Vector Machine. An advantage of the method proposed in this paper over the aforementioned state-of-the-art classifiers is the descriptive ability of the classifications it produces. The descriptive model can be of great help to aid a doctor in the diagnosis of Alzheimer's Disease, or even further the understand of how Alzheimer's Disease affects the hippocampus.

## I. INTRODUCTION

Alzheimer's Disease (AD) is a chronic neurodegenerative disease which is responsible for 60% to 70% cases of dementia and affects around 6% of the population who are aged 65 and above; the percentage affected by the disease increases with age [1]. Symptoms of AD based on the DSM-IV TR criteria include: memory impairment; aphasia (language disturbance); apraxia (impaired motor functionality); agnosia (failure to recognise objects) and a disturbance in executive functioning (the management of cognitive processes) [2]. Early detection of AD is challenging as there have been no biological markers found to definitively diagnose it at this stage. Current diagnosis of AD involves clinical approaches which are a set of neuropsychological tests to assess the patient [3], thus AD can only be diagnosed once it has had a significant effect on the patient's lifestyle. While there is no cure for an AD sufferer, symptomatic treatments exist to help patients reduce the symptoms they are suffering from. Clinical trials are still running to develop new treatments aimed to lower the chance of developing AD or delaying the onset and progression of it [4].

Automated classification of AD is the process where a subject is classified as a healthy control (HC) or suffering from AD. Most research in this area tends to use either neuropsychological tests as data; or more commonly, magnetic resonance images (MRIs). The classifier developed in this paper uses MRI data, thus MRI data will be the focus when investigating previous research. There are multiple types of MRIs, two of these types are structural MRIs and functional MRIs. Structural MRIs are the 3D image of a subject's brain recording the intensity at each point that is measured (the intensity refers to whether the part of the brain at the measured location is white matter, grey matter or cerebral spinal fluid). Functional MRIs include the structural MRI of the brain as well as measuring brain activity from changes in blood flow in the brain. There exist multiple challenges to the classification of MRI data, the first problem which occurs is artifacts produced by the MRI scanner, these include ghosting which is caused by the subject moving potentially due to respiratory motion and heartbeats [5]. Fixing these artifacts is necessary else the MRIs are likely to be of bad quality and may produce data which will lead to the creation of an inadequate classifier for AD. Another challenge is the large dimensionality of the data which leads to the curse of dimensionality where as the dimensions increase the data becomes more sparse and harder to find patterns.

[6] investigates feature selection methods (these methods are: recursive feature elimination, correlation filters, random forest filter) for SVMs applied to extracted data from the FreeSurfer processed MRIs. One-versus-one SVMs are combined with a voting mechanism to create a three-class classifier capable of classifying between HC, MCI and AD. Intra cranial volume normalisation (ICV) is also tested on the subjects to determine whether it is beneficial to the accuracy of the classifier or not. The best accuracy is achieved by the SVM which uses no ICV normalisation and the random forest filter. [7] also used extracted FreeSurfer features from MRI where features were selected based on a priori knowledge, and features were also combined with each other to reduce the dimensionality; a high classification accuracy of 73% was

TABLE I: Statistics of the data used in this paper

Diagnosis	Number of subjects	% Male	Age (mean $\pm$ std)
HC	145	45.5	73.9 $\pm$ 6.3
AD	142	54.2	75.0 $\pm$ 7.8

achieved. Previous work which uses Bayesian modelling in application to AD diagnosis includes [8] which uses Bayes' Theorem to calculate likelihood ratios for multiple tests which are used to diagnose a patient with AD. These likelihood ratios are then combined to find the tests which have the best joint predictive power on the diagnosis of the subjects. Authors in [9] propose a sparse Bayesian multi-task learning algorithm on MRI data extracted by FreeSurfer to deduce relationships between neuroimaging measures and cognitive scores to show how changes in the structure of a brain can affect the cognitive condition of it. [10] uses a Bayesian classifier to distinguish between HC and AD based on the clusters of voxel data of structural MRI scans.

## II. DATA PRE-PROCESSING

The data used in this paper was created using 287 MRI scans downloaded from the ADNI database and processed with FreeSurfer version 5.3 [11] with the optional command to segment the hippocampal subfields as well as perform the reconstruction of the entire brain. The hippocampal subfields consist of 16 volumetric measurements: the entire left hippocampus, the entire right hippocampus, the left and right presubiculum, the left and right subiculum, the left and right cornu ammonis (CA) 1, the left and right CA 2-3, the left and right dentate gyrus, the left and right fimbria, and the left and right hippocampal fissure. The criteria used to refine the search to find these 287 subjects was: that they were the baseline scan for each subject - this is the initial scan taken and initial diagnosis given to the subject; the slice thickness of scan was 1.5mm and it was weighted in T1; and that FreeSurfer was able to analyse the MRI (in rare cases FreeSurfer crashes during its execution if it is unable to process the MRI). A breakdown of the information of the subjects used in this study can be found in Figure I, which shows the distribution of attributes of the subjects such as diagnosis, age and gender.

ICV normalisation is a technique commonly used in classification of dementia from structural MRI; it involves transforming the volume measurements to account for different head sizes and total brain size as age and gender will affect the brain size which will in turn affect the volumes of the regions inside the brain. There are multiple methods to calculate ICV described in [12], the simplest of which is the proportional method  $v' = \frac{v}{ICV}$ , where  $v'$  is the ICV normalised value of the volume  $v$ . Another method where a linear relation exists between  $v$  and  $ICV$  such that  $v = b \times ICV + m$  where  $b$  is the gradient and  $m$  is the intercept. Then  $v' = b + \frac{m}{ICV}$ . A third method, the residual method also assumes a linear relation between  $v$  and  $ICV$  and transforms  $v$  by removing this linear relation,  $v' = v - b(ICV - \overline{ICV})$ .

In the data used in this paper there is not a linear relation between the hippocampal volumes and the ICV of the patient,

thus using ICV normalisation may cause loss of relationships between variables as two of the three methods previously described require a linear relationship between the volumes and ICV. Also ICV normalisation has been shown to lead to bias in volume measurements based on the age or gender of the subjects [13]. While there are multiple arguments against the usage of ICV normalisation on the data used in this paper, it will be verified as it may improve the accuracy of the method regardless of the potential downsides. The proportional method will be used,  $v' = \frac{v}{ICV}$ , to compute the ICV normalised volumes.

## III. PROBABILITY-BASED CLASSIFIER

### A. Linear Discriminant Analysis to Partition the Feature Space of a Single Region

Linear Discriminant Analysis (LDA) is a method to find the optimal linear hyperplane to separate two classes with the least margin of error. In this case it will be applied to single region to determine an optimal threshold where the values that lie below the threshold belong to one class, and the values that lie above the threshold belong to another class. Figure 1 shows the probability density functions for each of the hippocampal volumes for HC and AD patients and also the threshold for each attribute generated by LDA. From this it can be seen that generally the volumes for HC subjects are greater than the volumes than AD patients except in the left and right hippocampal fissures; however, the fissures are measurements of a gap rather than the volume of a mass in the hippocampus. It also shows the thresholds created when LDA is used on the attributes to classify the diagnosis of the subject. As the hippocampal fissure measurements do not provide any predictive power for the diagnosis of a subject these measurements will be ignored and only the other 14 attributes will be used.

Using the hyperplanes found during the LDA process, each subject's hippocampal attributes are tested to verify whether the value lies in the HC-sized side of the hyperplane or the AD-sized side. A value belongs to the HC-sized side if it lies above the threshold, and AD-sized otherwise. If the subject lies in the HC-sized side of the hyperplane then a -1 is recorded for that attribute, and for the AD-sized side, a 1 is recorded; this generates a binary feature vector for each subject as the attributes have been transformed from being a continuous value to having one of two possible values. An example of this process on a subject is shown in Table II. Intuitively if a higher count of 1 is found it should mean the subject is more likely to have AD as more of the hippocampal attributes indicate a likelihood that AD is present; and vice versa - if a higher count of -1 was found it means there is a likelihood of HC.

This is shown in Figure 2, where the bar chart showing the distribution of subjects with the number of positive attributes and their actual diagnosis. This bar chart shows a correlation between the number of positive attributes and a diagnosis of AD; and also a correlation between the number of negative attributes and HC subjects. However, there are subjects with a

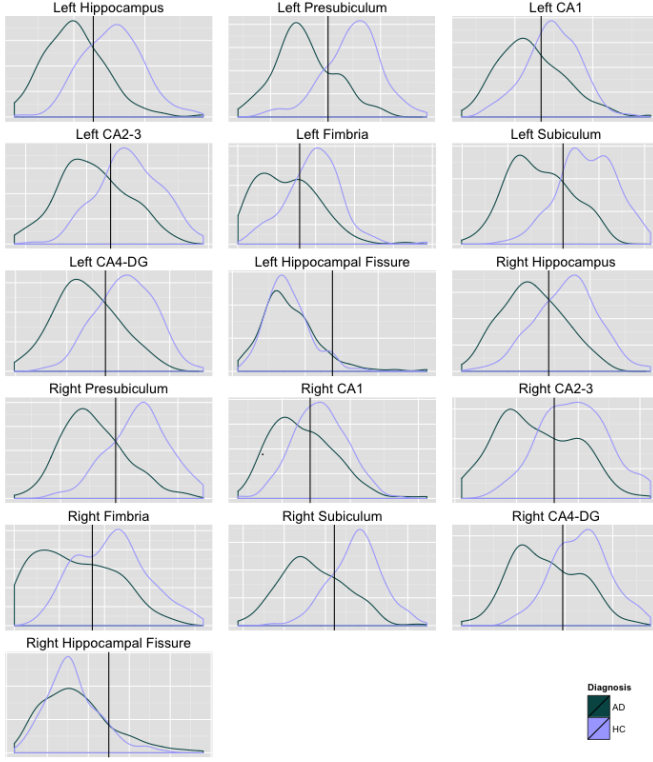


Fig. 1: Density graphs of the hippocampal volumes and thresholds found using LDA (the axis labels have been removed for clarity)

low number of positive attributes yet still classified with AD and vice versa for HC patients with a low number of negative attributes - in particular there are two HC subjects with all positive attributes; and two AD subjects with all negative attributes.

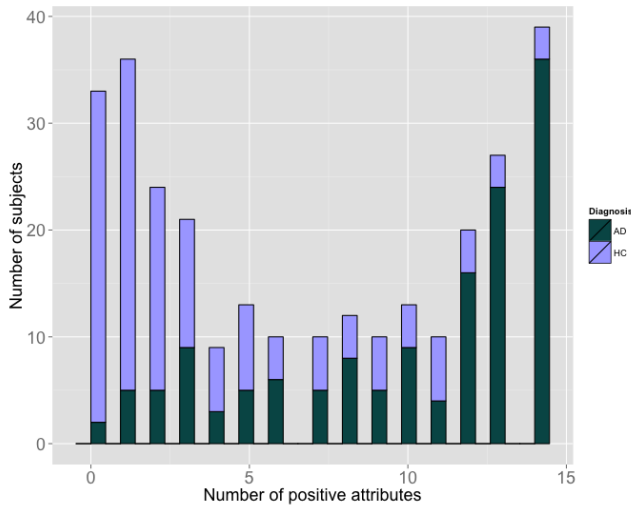


Fig. 2: Bar chart showing the number of positive attributes in subjects

TABLE II: LDA process for a subject with AD. A -1 in the 'Binary Feature' column means that region has been identified as a negative attribute, and a 1 means a positive attribute.

Attribute Name	Attribute Value	Binary Feature
Left Hippocampus	2353.56	-1
Left Presubiculum	3979.98	1
Left CA1	2101.93	1
Left CA2-3	5403.29	1
Left Fimbria	207.403	1
Left Subiculum	3979.98	1
Left CA4-DG	3257.18	1
Right Hippocampus	2350.65	1
Right Presubiculum	2433.25	1
Right CA1	2877.41	-1
Right CA2-3	6835.59	1
Right Fimbria	127.686	1
Right Subiculum	4262.02	-1
Right CA4-DG	3956.99	-1

#### IV. PROBABILITY WEIGHT-BASED CLASSIFICATION ALGORITHM

Using the LDA thresholds, an attribute is defined as being negative if  $x_n > t_n$ , where  $t_n$  is the  $n^{th}$  threshold value, and positive otherwise. We will define a function  $\phi(x_n)$  which will classify a single attribute as being negative or positive which will be given values of -1 and 1 respectively, the function is defined in Equation 1, where  $1_A$  is the indicator function and returns 1 if the conditional statement input to the function evaluates as true, and it returns 0 if the statement evaluates as false. Based on the number of positive attributes for a subject, a classification can be made as to whether the subject is HC or AD, as for an AD patient, the number of positive regions found would be expected to be greater than the number of negative regions found, this algorithm will be referred to as the Binary Region Classification algorithm (BRC). BRC works by summing the binary feature vector (which is comprised of -1 and 1 rather than 0 and 1) and if the value is above 0 a classification of HC is made; if it's 0 or below then an AD classification is made. BRC is shown in Equation 2.

$$\phi(x_i) = 1_A(x_i \leq t_i) - 1_A(x_i > t_i) \quad (1)$$

$$BRC(\vec{x}) = \sum_{i=1}^{dim(\vec{x})} \phi(x_i) \quad (2)$$

One downside to the BRC is that each region is given the same weight as every other region (1 or -1) - it assumes all regions have the same predictive power. This is not the case, some attributes will be better at identifying AD in a subject given the attribute is positive. This probability can be written as  $P(AD | x_i \leq t_i)$ . The probability of a patient being HC given the attribute is negative is given by  $P(HC | x_n > t_i)$ . The predictive powers for each volume measurement are shown for HC and AD in Table III and they can be used to weight the attributes: rather than each attribute having an integer value of -1 or 1. Table III also shows some additional interesting information: the top four predictors for both AD

and HC are the same (though in a slightly different order of ability for each diagnosis): Left Subiculum, Right Subiculum, Left Presubiculum and Right Presubiculum; this matches with the findings of [14] where it was found that atrophy of the subiculum and presubiculum were the best hippocampal markers for detection of AD.

TABLE III: Ability of single regions to diagnose a subject, to save space the table headers are denoted as  $x, y$  relating to the probability of a correct diagnosis of  $x$  given that the attribute is classed as  $y$ -sized. An attribute is HC-sized if it is negative, and AD-sized if it is positive.

Region Name	HC, HC	HC, AD	AD, HC	AD, AD
Left Hippocampus	0.75	0.25	0.25	0.75
Left Presubiculum	0.79	0.23	0.21	0.77
Left CA1	0.68	0.31	0.32	0.69
Left CA2-3	0.74	0.26	0.26	0.74
Left Fimbria	0.66	0.32	0.34	0.68
Left Subiculum	0.81	0.18	0.19	0.82
Left CA4-DG	0.74	0.26	0.26	0.74
Right Hippocampus	0.70	0.31	0.30	0.69
Right Presubiculum	0.80	0.22	0.20	0.78
Right CA1	0.60	0.40	0.40	0.60
Right CA2-3	0.73	0.28	0.27	0.72
Right Fimbria	0.67	0.34	0.33	0.66
Right Subiculum	0.76	0.20	0.24	0.80
Right CA4-DG	0.76	0.25	0.24	0.75

Using this weighting system which will be referred to as Probability-Weight Classification (PWC), two variants of the algorithm are proposed and tested.  $PWC_1(X)$  and  $PWC_2(X)$  are the variants and they adopt two equations which use the summation of the probability weights, and the resultant value is used to classify the subject as HC if it is less than 0 and AD otherwise. The variants,  $PWC_1(X)$  and  $PWC_2(X)$ , are shown in Equations 3 and 4 respectively.

Built-in feature selection can be implemented for these algorithms by having a threshold where if the predictive ability of a region falls below this threshold then it is discarded. The reasoning behind this is that regions which are worse at predicting HC or AD in a subject and thus the information they provide may lead to more misclassifications. Therefore if a thresholds  $P_T$  is provided such that all regions with a predictive power lower than  $P_T$  in the context of  $P(HC | x_i > t_i)$  and  $P(AD | x_i \leq t_i)$  are ignored (note that the minimum number of features that can remain are 1 no matter the value of  $P_T$ ).

#### A. ICV Normalisation

Two methods of ICV normalisation will be tested, the proportional method discussed earlier where  $v' = \frac{1}{ICV}v$  and a custom method which involves only normalising the volumes which lie outside the mean of all the subjects' values of that volume plus or minus the standard deviation of the values for the volumes. This is shown in Equation 5, where  $sd(ICV)$  is the standard deviation of the volume values and will be referred to as ICV normalisation SD. The ICV normalisation SD method was designed because in the initial testing of the algorithm, many of the false positives (HC misclassified as

AD) and false negatives (AD misclassified as HC) were from patients who had a very high or very low ICV in comparison to the other patients (such that their ICV was outside the range of the mean  $\pm$  the standard deviation of the ICVs of all subjects).

$$v' = \alpha v + (1 - \alpha) \frac{\overline{ICV}}{ICV} v$$

$$\alpha = 1_A \left( (\overline{ICV} - sd(ICV)) > ICV > (\overline{ICV} + sd(ICV)) \right) \quad (5)$$

#### B. Naive Bayes and Support Vector Machine Classification

This paper describes a custom method being used to classify the data, to evaluate the performance of this method it will be compared to two state-of-the-art classification methods: a Naive Bayes classifier and a Support Vector Machine (SVM). The Naive Bayes classifier is a probabilistic classifier which applies Bayes' theorem on a set of independent features, in this case the independent features are the hippocampal volume measurements for each subject. An SVM is a non-linear classifier using a kernel to transform data into a higher dimensional feature space and then classify the data in the new feature space. The kernels which will be tested in this paper are the linear kernel  $k(x, x') = \langle x, x' \rangle$  and the Gaussian radial basis filter (RBF) kernel  $k(x, x') = \exp(-\sigma \|x - x'\|^2)$ .

### V. EXPERIMENTAL SETUP

After the initial pre-processing of the data with FreeSurfer, it was imported into the data mining software KNIME [15] with the use of the library K-Surfer [16]. This allowed the statistical data output from FreeSurfer to be merged into a single table rather than multiple text-based files. Using the KNIME the structural MRI measurements were merged with the corresponding patient's details - their age, gender and diagnosis; the merged data was then saved in a single comma separated file. The algorithm was implemented and tested using the statistical programming language R [17] using the data output from KNIME as the input. External libraries for R provided the implementation for the LDA classifier [18]; and also the SVM [19] and the Naive Bayes [20] classifier which were used to compare the algorithm proposed in this paper.

### VI. RESULTS

Initial testing in Figures 3 and 4 show that the optimal probability thresholds are found between 0.7 and 0.8 (Figure 3), so those set the range for the probability thresholds which will be tested further. The results for the BRC algorithm are found in Table IV where BRC is used on data where ICV normalisation has not been applied; Table V shows BRC applied to ICV normalised data and Table VI is BRC applied to data which uses the ICV normalisation SD method. The results for the PWC algorithm are found in Table VII where  $PWC_1$  and  $PWC_2$  are used on data which hasn't been ICV normalised;  $PWC_1$  and  $PWC_2$  are applied to data which have been ICV normalised and the results are in Table VIII; and in Table IX,  $PWC_1$  and  $PWC_2$  have been applied to data

$$\text{PWC}_1(\vec{x}) = \sum_{i=1}^{\dim(\vec{x})} \left( \frac{\phi(x_i) - 1}{2} \left( P(HC|x_n > t_i) - P(AD|x_n > t_i) \right) + \frac{\phi(x_i) + 1}{2} \left( P(HC|x_n \leq t_i) - P(AD|x_n \leq t_i) \right) \right) \quad (3)$$

$$\text{PWC}_2(\vec{x}) = \sum_{i=1}^{\dim(\vec{x})} \left( \frac{\phi(x_i) - 1}{2} P(HC | x_n > t_i) - \frac{\phi(x_i) + 1}{2} P(AD | x_i \leq t_i) \right) \quad (4)$$

TABLE IV: BRC without ICV normalisation and 10-fold cross validation repeated ten times

$P_T$	Acc	Sen	Spe
<b>0.7</b>	0.785	0.777	0.792
<b>0.71</b>	<b>0.795</b>	0.795	0.795
<b>0.72</b>	<b>0.795</b>	0.8	0.79
<b>0.73</b>	0.793	0.799	0.787
<b>0.74</b>	0.789	0.798	0.781
<b>0.75</b>	0.782	0.792	0.774
<b>0.76</b>	0.784	0.787	0.781
<b>0.77</b>	0.775	0.792	0.758
<b>0.78</b>	0.765	0.781	0.749
<b>0.79</b>	0.695	0.613	0.775
<b>0.8</b>	0.606	0.331	0.876

with the ICV normalisation SD method used. A Naive Bayes classifier has been tested on the data in Table XI as it, like PWC, also uses probability to classify data so they both share a similarity. SVMs were also applied to the data in Table X because out of various classifiers applied to the data, SVMs outperformed them all thus it would be useful to compare the classifiers developed in this paper to what might be one of the highest accuracies achievable on the data with commonly used classifiers.

For all the results 10-fold cross validation was repeated ten times using a different set of folds for each repetition. All of the classifiers tested used the same sets of folds so they were training and testing on the same data sets. 10-fold cross validation is a form of k-fold cross validation [21] and works by splitting the data set into k folds where each fold contains an equal (or near equal) number of samples. Next, k iterations of training and testing are performed where  $k - 1$  folds are used for training and the leftover fold is used for testing. Eventually all folds will have been tested (using a different  $k - 1$  folds for training at each iteration) and then an accuracy can be computed between the actual class of all of the samples of the data and the predicted class for all the data samples. The 10-fold cross validation is repeated ten times for all classifiers in order to try and eliminate any bias towards the data set being tested.

## VII. DISCUSSION

The highest accuracy achieved was by BRC with an accuracy of 85.8% using a probability threshold of 0.78 on ICV normalised data. This outperformed even the SVM which obtained an accuracy of 84.9% using a linear kernel on ICV normalised data and also outperformed the Naive Bayes

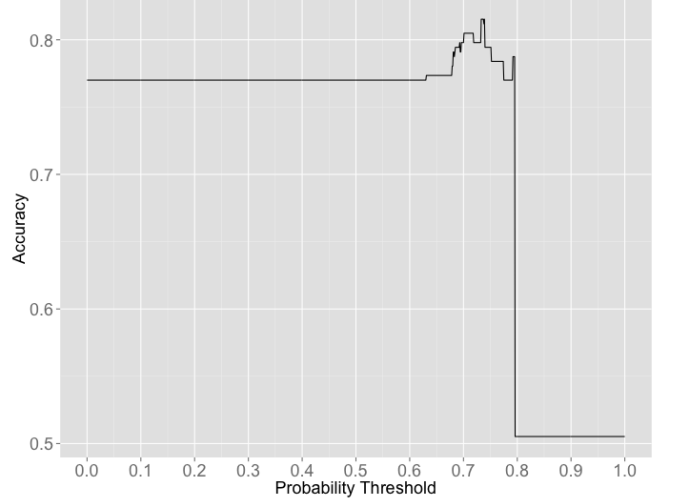


Fig. 3: Accuracy achieved via  $PWC_1$  on the non-ICV normalised entire dataset as the probability threshold is changed

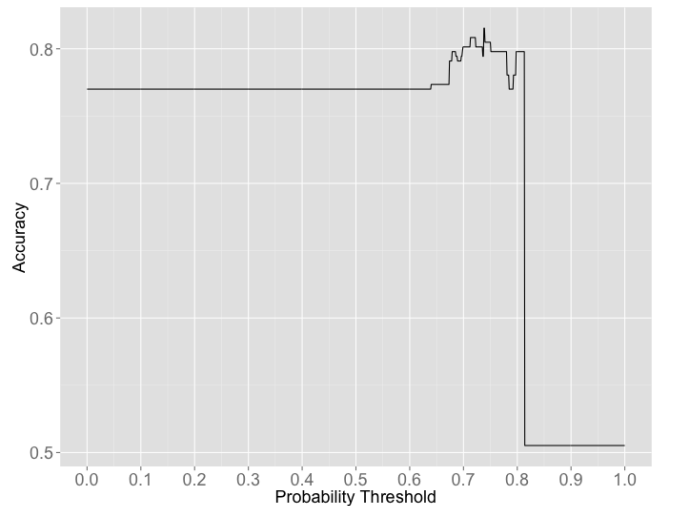


Fig. 4: Accuracy achieved via  $PWC_1$  on the ICV normalised entire dataset as the probability threshold is changed

TABLE V: BRC with ICV normalisation and 10-fold cross validation repeated ten times

$P_T$	Acc	Sen	Spe
<b>0.7</b>	0.835	0.805	0.863
<b>0.71</b>	0.836	0.804	0.867
<b>0.72</b>	0.838	0.803	0.872
<b>0.73</b>	0.84	0.801	0.879
<b>0.74</b>	0.838	0.799	0.876
<b>0.75</b>	0.832	0.792	0.872
<b>0.76</b>	0.838	0.8	0.874
<b>0.77</b>	0.849	0.813	0.883
<b>0.78</b>	<b>0.858</b>	0.832	0.883
<b>0.79</b>	0.852	0.837	0.868
<b>0.8</b>	0.832	0.825	0.839

TABLE VI: *BRC* with ICV normalisation SD with 10-fold cross validation repeated ten times

$P_T$	Acc	Sen	Spe
<b>0.7</b>	0.841	0.821	0.86
<b>0.71</b>	0.842	0.82	0.862
<b>0.72</b>	0.843	0.818	0.867
<b>0.73</b>	0.844	0.812	0.874
<b>0.74</b>	0.84	0.813	0.866
<b>0.75</b>	0.839	0.815	0.862
<b>0.76</b>	0.842	0.827	0.857
<b>0.77</b>	<b>0.846</b>	0.841	0.85
<b>0.78</b>	<b>0.846</b>	0.849	0.843
<b>0.79</b>	0.843	0.851	0.835
<b>0.8</b>	0.837	0.84	0.833

TABLE VII:  $PWC_1$  and  $PWC_2$  with no ICV normalisation with 10-fold cross validation repeated ten times

$P_T$	No ICV Normalisation					
	$PWC_1$			$PWC_2$		
	Acc	Sen	Spe	Acc	Sen	Spe
<b>0.7</b>	0.785	0.777	0.792	0.786	0.779	0.792
<b>0.71</b>	<b>0.795</b>	0.795	0.795	0.795	0.796	0.795
<b>0.72</b>	<b>0.795</b>	0.8	0.79	0.795	0.8	0.79
<b>0.73</b>	0.793	0.799	0.787	0.793	0.799	0.787
<b>0.74</b>	0.789	0.798	0.781	0.789	0.798	0.781
<b>0.75</b>	0.782	0.792	0.774	0.782	0.791	0.774
<b>0.76</b>	0.784	0.787	0.781	0.783	0.786	0.781
<b>0.77</b>	0.775	0.792	0.758	0.775	0.792	0.758
<b>0.78</b>	0.765	0.781	0.749	0.765	0.781	0.749
<b>0.79</b>	0.695	0.613	0.775	0.695	0.613	0.775
<b>0.8</b>	0.606	0.331	0.876	0.606	0.331	0.876

TABLE VIII:  $PWC_1$  and  $PWC_2$  with ICV normalisation with 10-fold cross validation repeated ten times

$P_T$	ICV Normalisation					
	$PWC_1$			$PWC_2$		
	Acc	Sen	Spe	Acc	Sen	Spe
<b>0.7</b>	0.828	0.812	0.845	0.841	0.821	0.86
<b>0.71</b>	0.831	0.816	0.845	0.842	0.82	0.862
<b>0.72</b>	0.832	0.815	0.848	0.843	0.818	0.867
<b>0.73</b>	0.835	0.818	0.85	0.844	0.812	0.874
<b>0.74</b>	0.838	0.823	0.854	0.84	0.813	0.866
<b>0.75</b>	0.835	0.815	0.854	0.839	0.815	0.862
<b>0.76</b>	0.839	0.823	0.854	0.842	0.827	0.857
<b>0.77</b>	0.847	0.837	0.857	0.846	0.841	0.85
<b>0.78</b>	<b>0.856</b>	0.854	0.859	0.846	0.849	0.843
<b>0.79</b>	0.85	0.84	0.859	0.843	0.851	0.835
<b>0.8</b>	0.833	0.799	0.866	0.837	0.84	0.833

TABLE IX:  $PWC_1$  and  $PWC_2$  with ICV normalisation SD with 10-fold cross validation repeated ten times

$P_T$	ICV Normalisation SD					
	$PWC_1$			$PWC_2$		
	Acc	Sen	Spe	Acc	Sen	Spe
<b>0.7</b>	0.829	0.813	0.845	0.843	0.823	0.862
<b>0.71</b>	0.83	0.815	0.845	0.842	0.822	0.862
<b>0.72</b>	0.831	0.814	0.848	0.843	0.818	0.867
<b>0.73</b>	0.834	0.818	0.85	0.844	0.812	0.874
<b>0.74</b>	0.838	0.823	0.854	0.84	0.813	0.866
<b>0.75</b>	0.834	0.811	0.857	0.838	0.816	0.859
<b>0.76</b>	0.84	0.824	0.856	0.842	0.827	0.857
<b>0.77</b>	0.847	0.837	0.857	0.846	0.841	0.85
<b>0.78</b>	<b>0.856</b>	0.854	0.859	0.846	0.849	0.843
<b>0.79</b>	0.85	0.84	0.859	0.843	0.851	0.835
<b>0.8</b>	0.832	0.799	0.865	0.834	0.842	0.827

TABLE X: SVM with 10-fold cross validation repeated ten times

	Kernel	Acc	Sen	Spe
<b>SVM without ICV normalisation</b>	Linear	0.806	0.815	0.797
	RBF	0.823	0.831	0.814
<b>SVM with ICV normalisation</b>	Linear	<b>0.849</b>	0.847	0.852
	RBF	0.835	0.85	0.819
<b>SVM with ICV normalisation SD</b>	Linear	0.835	0.837	0.833
	RBF	0.841	0.866	0.815

classifier with a maximum accuracy of 84.7%. The difference between BRC and PWC was negligible with PWC having a maximum accuracy of 85.6% (compared to BRC's 85.8%) thus showing that there is no need to complicate the method using the probability-based weights that is used by PWC and instead use the binary weights from BRC.  $PWC_1$  always achieved a higher maximum accuracy than  $PWC_2$ , however, at lower probability thresholds (0.7 - 0.75),  $PWC_2$  performs better than  $PWC_1$  so it is better able to cope with less relevant features.

ICV normalised data far outperformed non-ICV normalised data; in BRC and PWC using ICV normalisation improved the maximum accuracy by over 5% in both cases, and all the accuracies at each probability threshold used were improved. With the Naive Bayes classifier and the SVM, ICV normalisation also improved the results, with the SVM being the classifier able to cope with the non-ICV normalised data the best by achieved the highest accuracy of 82.3% on it. The ICV normalisation SD method described in this paper in all cases par one performs worse than standard ICV normalisation, and it always outperforms ICV normalisation. For the Naive Bayes classifier however, ICV normalisation SD outperforms ICV normalisation by 1.5%, though this is only a marginal amount and since ICV normalisation SD was only better than standard

TABLE XI: Naive Bayes with 10-fold cross validation repeated ten times

Data Transform Method	Acc	Sen	Spe
<b>NB without ICV normalisation</b>	0.777	0.732	0.821
<b>NB with ICV normalisation</b>	0.833	0.81	0.855
<b>NB with ICV normalisation SD</b>	<b>0.847</b>	0.817	0.876



ICV normalisation in one case, the SD method isn't worth using.

The descriptive model produced by this algorithm is shown in Table II, each region of a subject's hippocampus is given a binary feature of 1 or -1 depending on the value of the volume measurement for that region of the hippocampus. If a subject has more negative attributes than positive attributes then a diagnosis of HC is made; and more positive attributes than negative attributes then an AD diagnosis is made. There is a case which could occur for certain subjects where the number of negative attributes is equal to the number of positive attributes; in this case, a classification of AD will be made - the reasoning behind this is that it would be better (regarding the patient's interest) for a healthy patient to receive treatment for AD rather than an AD patient to go ignored. An alternative would be to not make a decision and leave the diagnosis unknown which could potentially lead to doctor's monitoring the development of the patient's brain over the foreseeable future.

### VIII. CONCLUSION

The work in this paper has created a classifier which outperforms an SVM and a Naive Bayes classifier at accuracy in diagnosing a patient as HC or AD based on a baseline structural MRI scan. As well as attaining a higher accuracy it can extract a easily understood description of why a diagnosis was made, whereas an SVM doesn't have this advantage as an SVM works by transforming the data into a higher dimensional space where descriptive information about the data is lost. The descriptiveness of this model could aid a doctor by pointing to regions in the hippocampus that the doctor should look at - if we let all negative features be a warning sign of AD, then say the left presubiculum is negative, the classifier could advise the doctor to manually check the left presubiculum of the subject's MRI to see if there is anything about that region that could correlate with AD. This work has also shown that the best features in the hippocampus are the left presubiculum, right presubiculum, left subiculum and right subiculum matching with current literature on this topic [14].

There is also potential for the algorithm to have a parallel execution in both the training and predicting parts of it. The training of the algorithm could be parallelised by: creating the thresholds for the decision boundary in parallel and evaluating different probability thresholds in parallel. Regarding the prediction part of the algorithm: the proposed methods in this paper are based on the summation of weights and each of the weights are independent of the other weights, this means that the weights can be computed in parallel.

Future work involves applying this algorithm to the three class HC, AD and also Mild Cognitive Impairment sufferers and seeing if it performs as well. After the thresholds are created, region are given the diagnosis of negative or positive which converts a continuous variable into a binary variable, rule mining could be applied to the binary variables to generate rule associations to further the understanding of AD in the hippocampus. In this paper only the hippocampal subfields are

used, the classifier's performance when applied to the cortical and subcortical fields of the brain could improve so it is worth seeing how well it works applied to the additional features. Another idea is to use subspace clustering to find a subspace where the subjects can be grouped into clusters and see if rules can be deduced from the clusters that are able to be merged with the rules found by this classifier to create a better classifier.

### ACKNOWLEDGMENT

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

### REFERENCES

- [1] A. Burns and S. Iliffe, "Alzheimer's disease," *BMJ*, vol. 338, 2009.
- [2] H. H. Feldman, *Atlas of Alzheimer's Disease*. Informa Healthcare, 2007.
- [3] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939-944, Jul 1984.
- [4] H.-W. Klafki, M. Staufenbiel, J. Kornhuber, and J. Wiltfang, "Therapeutic approaches to alzheimer's disease," *Brain*, vol. 129, no. 11, pp. 2840-2855, 2006.
- [5] T. B. Smith and K. S. Nayak, "Mri artifacts and correction strategies," *Imaging in Medicine*, vol. 2, no. 4, pp. 445-457, 2010.
- [6] A. Sarica, G. D. Fatta, G. Smith, M. Cannataro, and J. D. Saddy, "Advanced feature selection in multinomial dementia classification from structural mri data," *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pp. 82-91, 2014.
- [7] L. Sørensen, A. Pai, C. Anker, I. Balas, M. Lillholm, C. Igel, and M. Nielsen, "Dementia diagnosis using mri cortical thickness, shape, texture, and volumetry," *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pp. 111-118, 2014.
- [8] M. J. Prince, "Predicting the onset of alzheimer's disease using bayes' theorem," *American Journal of Epidemiology*, vol. 143, no. 3, pp. 301-308, 1996. [Online]. Available: <http://aje.oxfordjournals.org/content/143/3/301.abstract>
- [9] J. Wan, Z. Zhang, J. Yan, T. Li, B. Rao, S. Fang, S. Kim, S. Risacher, A. Saykin, and L. Shen, "Sparse bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in alzheimer's disease," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 940-947.
- [10] C. Plant, S. J. Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda, A. W. Bokde, H. Hampel, and M. Ewers, "Automated detection of brain atrophy patterns based on {MRI} for the prediction of alzheimer's disease," *NeuroImage*, vol. 50, no. 1, pp. 162 - 174, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811909012312>
- [11] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, pp. 341-355, 2002.

- [12] R. Nordenskjöld, F. Malmberg, E.-M. Larsson, A. Simmons, H. Ahlström, L. Johansson, and J. Kullberg, "Intracranial volume normalization methods: Considerations when investigating gender differences in regional brain volume," *Psychiatry Research: Neuroimaging*, vol. 231, no. 3, pp. 227 – 235, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925492714003084>
- [13] R. Nordenskjöld, F. Malmberg, E.-M. Larsson, A. Simmons, S. J. Brooks, L. Lind, H. Ahlström, L. Johansson, and J. Kullberg, "Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements," *NeuroImage*, vol. 83, pp. 355 – 360, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811913007064>
- [14] G. A. Carlesimo, F. Piras, M. D. Orfei, M. Iorio, C. Caltagirone, and G. Spalletta, "Atrophy of presubiculum and subiculum is the earliest hippocampal anatomical marker of alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 24 – 32, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2352872915000044>
- [15] M. R. Berthold, N. Cebon, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel, "KNIME: The Konstanz Information Miner," in *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007.
- [16] A. Sarica, G. Di Fatta, and M. Cannataro, "K-surfer: A knime-based tool for the management and analysis of human brain mri freesurfer/fsl data," *Frontiers in Neuroinformatics*, no. 3, 2014. [Online]. Available: <http://www.frontiersin.org/neuroinformatics/10.3389/conf.fninf.2014.18.00003/full>
- [17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org/>
- [18] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, ISBN 0-387-95457-0. [Online]. Available: <http://www.stats.ox.ac.uk/pub/MASS4>
- [19] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – an S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004. [Online]. Available: <http://www.jstatsoft.org/v11/i09/>
- [20] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2014, r package version 1.6-4. [Online]. Available: <http://CRAN.R-project.org/package=e1071>
- [21] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, L. LIU and M. ZSU, Eds. Springer US, 2009, pp. 532–538. [Online]. Available: [http://dx.doi.org/10.1007/978-0-387-39940-9\\_565](http://dx.doi.org/10.1007/978-0-387-39940-9_565)